



International Journal for Innovative Engineering and Management Research

A Peer Reviewed Open Access International Journal

www.ijiemr.org

COPY RIGHT



ELSEVIER
SSRN

2022 IJIEMR. Personal use of this material is permitted. Permission from IJIEMR must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. No Reprint should be done to this paper, all copy right is authenticated to Paper Authors

IJIEMR Transactions, online available on 26th Apr 2022. Link

[:http://www.ijiemr.org/downloads.php?vol=Volume-11&issue= Spl Issue 04](http://www.ijiemr.org/downloads.php?vol=Volume-11&issue= Spl Issue 04)

DOI: 10.48047/IJIEMR/V11/SPL ISSUE 04/01

Title ASSOCIATION RULE BASED BOOK RECOMMENDATION SYSTEM USING NLP

Volume 11, SPL ISSUE 04, Pages: 1-12

Paper Authors

**Mrs. V. Sowjanya, M.Sai Sudha, 3 Sk.Sameena Parveen, P.Chinmaya Srinivas,
B.Hema Latha**



USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per **UGC Guidelines** We Are Providing A Electronic Bar Code

ASSOCIATION RULE BASED BOOK RECOMMENDATION SYSTEM USING NLP

¹Mrs. V. Sowjanya, M.Tech, (Ph.D.),

²M.Sai Sudha, ³Sk.Sameena Parveen, ⁴P.Chinmaya Srinivas, ⁵B.Hema Latha

¹Associate Professor, CSE, PSCMR CET, Vijayawada, Andhra Pradesh

^{2,3,4,5} Student, CSE, PSCMR CET, Vijayawada, Andhra Pradesh

¹sowjanya@pscmr.ac.in, ²mondisaisudha@gmail.com, ³simmi.neelu2001@gmail.com, ⁴chinmayasrinivas145@gmail.com, ⁵bevarahemalatha225@gmail.com

Abstract

Every year the university library introduces a many books. A lot of time needs to be spent by user and similarly, all the books are may or may not be effective. So, this results in wasting time. When the library contains more books, it is difficult for users to choose an appropriate book based on a particular subject. To overcome this situation, the unique and simple solution called as a Book Recommendation System. To implement this, the book recommendation system follows the step-by-step process. Firstly, It follows the collaborative filtering method. That means, It gathers the information from previous user's review to recommend the books based on their similar predilections. Gathering of the previous user's reviews of the different textbooks through the questionnaire form is the foremost task. To build a strong recommendation system, it considers both comments and ratings. NLP is the best choice for analyzing comments like whether the comment is given by the user is positive or negative or neutral. After this, the technique called as association rule mining is very helpful for strong recommendations. In this technique, the apriori calculates three parameters like support, confidence, and lift for each unique textbook based on comments and ratings. Finally, it provides the recommended data-set. This data-set contains only positive records and these are useful for the users. Based on the user requirements this system suggests textbooks for the users.

Keywords: Collaborative Filtering, NLP, Association Rule Mining, Apriori Algorithm, Logistic Regression, SVM, Decision Tree Classifier, Logistic Regression, SVM, Random Forest Regression.

I. INTRODUCTION:

Currently recommendation approaches are used to suggest the different items to the users based on their predilections of . The information filtering systems are also known as recommendation systems that are used to forecast the preferences or ratings of a user wants to give

a particular item. YouTube, Amazon, Netflix etc., are some of the examples of the

recommendation systems. The main usage of a recommendation is to, it can reduces the time for searching or finding a particular item. There exist many drawbacks for traditional method called as searching a perfect book based on the particular subject

for a student. For instance, let us consider a student, who is new to the programming language and picks up on advanced c language books, it will be very difficult for her to understand. Because without basic knowledge she couldn't understand the advanced programming level. As a result she can't read any textbook or there is a wasting of time for searching a book. In this case, the Library Book Recommendation System will help her to pick a perfect book based on her knowledge or book level.

To implement this system, firstly it collects the information from different users (like faculty, and students) through the direct survey (Questionnaire). This action called as collaborative filtering approach. All this collected information is used as a data-set in this project. This data-set contains different features such as subject name, textbooks related to the selected subjects, reader details, purpose of the reading the book, number of times the book was read, time taken to complete the book, book level, suggestion of book, rating of the book, comments of the book.

Here, the primary aim of this proposed system is to suggest the textbooks based on high ratings and positive comments. For this, it follows the below process. Firstly, NLP (Natural Language Processing) is used for the comment analysis. In this, it uses the Vader Sentiment technique from the Sentiment Intensity Analyzer and this algorithm gives the given comment as positive or negative, or neutral. But here, this system considering both comments and ratings. So, the apriori algorithm is used to find out the parameters like support, confidence, and lift. It suggest the textbooks based on high confidence, high support, and high lift values of the different textbooks

based on the comments. These suggested records are set to be as a final data-set.

Finally, the result recommended data-set is used to connect to a web based application for the book recommendation system of library. This web application can access any user on the internet. This application provides the user's required inputs and based on the user's inputs it suggests the textbooks. This application takes the subject name as the first input and then it provides two options like, In what way to search the book such as book level or understanding level of user. Based on this user inputs, it suggest the best textbooks for the users.

II. RELATED WORK:

In [1] A.Anoop et.al, designed a book recommendation system for library using collaborative filtering algorithm in cloud. This system has been implemented to make out the daily library work in simple manner without involving manual work. Here they developed a website and they provided a sign-in option for the user using an email-id followed by password. The book is recommended to the particular user based on his/her profile, search activity, and rating of the book. They developed an admin module, user module, and live chat module using the cloud to maintain and store this huge amount of data. The process of this collaborative filtering algorithm is to recommend the particular book and that book is frequently read by other users previously. So, every user gets the books which are top rated recommended books and which results in not wasting the time for searching a book. Based on the quality and rating of the book they have used an item-based technique. But this method was identified as reliable itself when compared

to user based collaborative filtering. Because the average item can get high ratings than the average user. They have also used an Apriori algorithm based on searching for a relevant book. This apriori algorithm is used for transaction those are depending on the items those are frequently visited and it produces the items based on the lead. The main aim of this existing system is to recommend the high rated books to the new users in an efficient way through the internet without any manual work. The main disadvantage is, we have to enter the book name correctly. Otherwise, it gives an error message and these systems should be secured.

In [2] Dhiman Sarma et.al, implemented a “Personalized Book Recommendation System using ML Algorithm”. Due to the covid-19 pandemic situation, there is a vast increase in online books. It is difficult to select a book from a large number of options. We offer a book recommendation system based on the clustering technique to address these issues. Generally, the collaborative filtered algorithm is used for the recommendation system, but the accuracy for this method is 88% which is low. We have another method known as content-based, which is used to recommend books based on similarities among users. But this has a major drawback that is, it ignores the current users’ rating while suggesting a new book. In this base paper, they used the clustering technique. In this technique, they have used k-means and cosine function algorithms for recommending books. The first step in K-means clustering is to collect the number of k-cluster books. The cosine function is also used to find the value between two non-zero vectors. The Euclidean dot product can be used to

calculate this. This technique has high accuracy. The accuracy can be calculated using ROC (Receiver Operation Characteristic). Between sensitivity and specificity, the ROC curve is displayed. The average sensitivity and specificity are 49.67% and 56.74%, respectively. Finally, a clustering method was utilised in this study to improve the accuracy of the book recommendation system.

In [3] Yonghong Tian et.al, proposed a hybrid algorithm based college library personalized recommendation system, to identify the problem, When the library is providing a huge number of books then the users are in the confusion to selecting a book among similar kinds of books. Based on a hybrid recommendation algorithm, the personalized recommendation system is intended for college libraries. This algorithm involves the applications of both content-based and collaborating recommendation algorithms in the recommendation of books based on the similarity between the users. A hybrid recommendation system combines both content-based and collaborative filtering. It first does both content-based and collaborative-based filtering individually, then combines them. Add content-based features to the collaborative filtering technique next. Finally, techniques are being unified into a single model. Before determining similarity, this article uses K-means clustering. Clustering is the problem of grouping a set of things so that objects in the same cluster are more similar than those in other clusters. Finally, the Spark big data computing platform is memory-based, with less reads and writes to hard drives due to the parallelization algorithm's high execution speed.

In [4] Avi Rana et.al, explored collaborative filtering with jaccard similarity algorithm for online book recommendation system. A recommendation system is a piece of software that proposes comparable things to consumers based on their interests or previous purchases. Nowadays, most e-commerce businesses use these techniques to entice consumers by providing what the buyer is likely to desire. The difficulties they confront include determining a priority and making appropriate suggestions. Collaborative filtering is used to build lists of things that are comparable to the buyer's tastes. However, it poses issues such as scalability, sparsity, and cold start. As a result, this research provides a recommendation based on collaborative filtering and Jaccard similarity. This will result in more accurate findings. For item-based suggestions, Jaccard similarity is employed. It computes the similarity between the two publications by adding the number of subscribers who rated both books in the numerator to the number of users who rated either of the two books in the denominator. It does not take into account the actual rating provided to a book by the user, but rather the number of people who have rated the books. The book was evaluated using Root Mean Square Error (RMSE) statistical accuracy criteria in this article. The RMSE quantifies the difference between the projected value and the actual user rating.

In [5] Sirikayon et.al, described a pre-processing matrix factorization on collaborative filtering algorithm based book recommendation system for library. Nowadays, recommendation algorithms are used to propose user-specific things based on their preferences. As a result, many

academic libraries strive to create effective and efficient book recommendation systems. For the process of book selection, this investigation used one of the most extensively utilized techniques known as collaborative filtering. Data sparseness is the limitation overcome by collaborative filtering. For this, we adopt the bios matrix factorization technique to solve the sparseness problem. The book recommendation is done by using the existing book records with help of timestamps. The performance ratings are based on accurate measurements and student satisfaction. The Pearson correlation and cosine similarity are used to calculate similarity and distance for user-based collaborative filtering. Prediction is based on the top-k users who are most similar to the user and their rating ratings. This forecast brings up the issue of data sparsity. Matrix factorization is an effective method to handle the above data sparsity problem

All these systems are suffer from cold start problem and some of the systems are developed based on only comments or based on only reviews. This results may not be give the correct output.

III.PROPOSED SYSTEM:

Here we recognize that, the previous users information is very useful for recommend the textbooks to the new users. So, the first step is considered as data collection through the questionnaire form. Internally, the data is stored from this form. Also, we are using NLP for comment analysis and Apriori algorithm for providing strong and good textbooks for the users. The proposed system follows both positive comments and high ratings of the textbooks for the desired output. Figure 1 depicts the general block diagram of the suggested architecture.

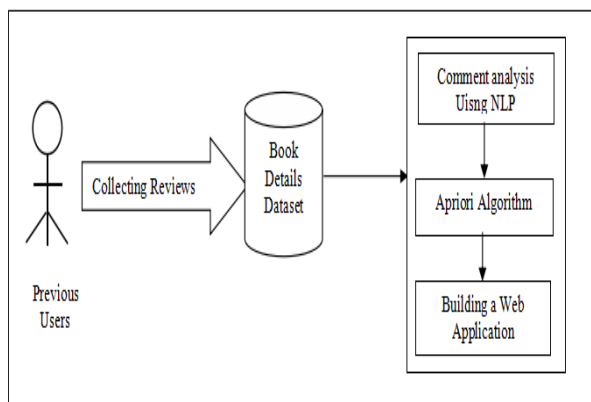


Figure 1: System Architecture for Library Book Recommendation System

To develop Library Book Recommendation System, we have used three steps:

1. Data Collection.
2. Comment Analysis Using NLP.
3. Apriori Algorithm For Association Rule Mining.

A. Data Collection

The foremost task is previous users reviews is known as Collaborative filtering. So, by collecting each textbook information from different users, the questionnaire form will helps. In this form contains different attributes like subject name, it's related text book, reader details, purpose of reading, no.of times read, duration to complete the book, book level, understanding level, rating of the book, comments. All these attributes are filled by previous users and these are not null. This collected data contains totally 984 records and 10 attributes. This data is represented as Book Details Data-set. Based on the selected subject and related its text book, the remaining attributed are filled. The below table represents the collected information from users.

Table 1: Book Details Dataset

| Subject Name | Text Book | Reader Details | ... | Book Level | Understanding Level | Ratings | Comments |
|--------------------|--|----------------|-----|--------------------|----------------------|---------|--|
| Maths-1 | N.P.Bali Engineering Mathematics, Lakshmi Publications | Faculty | ... | Basic Level | Slow Learner | 4 | Helpful and informative |
| Distributed System | Distributed Systems Principles- Tanenbaum PHI | Student | ... | Intermediate Level | Medium Paced Learner | 1 | Not suitable for academic purpose waste of time reading this book. |
| English-II | English Encounters Published by Maruthi Publishers. | Student | ... | Advanced Level | Medium Paced Learner | 3 | Somewhat tough to read |
| ... | ... | ... | ... | ... | ... | ... | ... |

Here, there is a conflict among comments and ratings. For example, a negative comment has high ratings and positive comment has low ratings. There is another problem that some comments has both positive words and negative words. So, it is difficult to predict the positive textbooks based on comments and ratings. To analyze comments, it follows NLP technique.

B. Comment Analysis Using NLP

Natural Language Processing is a technique for communicating between computers and humans. Humans can communicate in their own language. NLP allows computers to read text, hear voices,

assess sentiment, and determine which areas are important. This method of examining comments use natural language processing (NLP). Comments are analysis by sentiment analysis technique. Machine learning and natural language processing are used in sentiment analysis to determine if a sentence is negative, positive, or neutral. Natural Language Toolkit (nlTK) is a Python module that may be used for NLP in this procedure. A large portion of the data you may be examining is unstructured and comprises human-readable text. VADER (Valence Aware Dictionary for Sentiment Reasoning) is a text sentiment analysis model that is sensitive to both polarity (positive/negative) and intensity (strong) of emotion in this nltk package.

Here, the comments are taken as input and it calculates the score for the each comment using Vader Sentiment Model from sentiment Intensity analyzer. The process of analyzing the score is, First we are removing the 0-9 numbers in the comments and removing all the punctuation's like comma, full-stop, etc. Stop words are a group of terms that are widely used in a language. Stop words in English include 'the', 'are', 'is', 'a', and others. Stop words are frequently used in Text Mining and Natural Language Processing (NLP) to exclude words that are so frequently used that they contain very little meaningful information. The comment contains both positive and negative words. The score is calculated based on these words. It is computed as shown in equation 1.

$$\text{Score for comment analysis} = \frac{\text{Count of positive words}}{\text{Count of negative words} + 1} - (1)$$

Since there is no difference of values involved, the sentiment score value will always be more than 0. Also, adding 1 in the denominator would save from Zero Division Error. Based on the score, the comment is recognized as whether the given comment is positive, or negative, or neutral.

Table 2: Analyzing the comments using NLP

| Comments | Clean_review_text | Score | Score_Status |
|--|---|---------|--------------|
| Helpful and informative | helpful informative | 0.4215 | positive |
| Not suitable for academic purpose waste of time reading this book. | suitable academic purpose waste time reading book | -0.4215 | negative |
| Somewhat tough to read | somewhat tough read | -0.0534 | negative |
| Read 20 times then you can understand it. | Read times understand | 0 | neutral |
| | | | |

The following technique is used to analyzing the comment:

- If score ≥ 0.05 then Score_Status is "positive".
- Elif score < 0.05 then Score_Status is "negative".
- Elif score $== 0$ then Score_Status is "neutral".

By using the above formulas, we can determine whether the comment is positive, or negative, or neutral. But, there occurs a problem that, we cannot predict the correct output by taking previous users comments. Because, every user has their own perspective. For example, User A likes one textbook but another User B does not likes

the same textbook. So, we cannot determine good textbooks are recommended for which user. These are all the different problems are occurs. So, to overcome this situation, we implemented a Association Rule Mining. It can overcomes these problems and gives the strong recommendation for the new users.

C. Apriori Algorithm

Association rule mining is a data mining approach for uncovering the rules that govern associations and causal objects between groups of items. Association rule learning is a type of unsupervised learning method that recognizes a data item's dependency on another and maps it accordingly to make it more profitable. It allows for the finding of fascinating connections and relationships between large collections of data items. It comprises searching a database for patterns or co-occurrences using machine learning models. It finds frequent if-then connections, which are the rules of the association. An association rule consists of two parts: an antecedent (if) and a result (if) (then). Because the data is not labeled, an algorithm attempts to learn without the help of a teacher in unsupervised learning. The association rule is a descriptive rather than predictive method that is frequently used to identify noteworthy associations hidden in large datasets. One of the techniques used in Association Rule Learning for transaction data is the Apriori Algorithm. It allows us to mine the often recurring item-set to establish association rules between them. For instance, a list of items purchased by clients, information about frequently visited websites, and so forth. We can determine the values of three parameters using this a priori approach: support, confidence, and lift.

Support:

This indicates the popularity of an itemset as measured by the proportion of transactions in which it occurs. It is computed as shown in equation 2.

If x => y then

$$Support = \frac{frequency(x,y)}{Total\ no.\ of\ transactions} \quad - (2)$$

Confidence:

This indicates how probable it is that item Y will be purchased when item X is purchased. This is determined by the fraction of transactions that include item X and item Y. It is computed as shown in equation 3.

If x => y then

$$Confidence = \frac{frequency(x,y)}{frequency(x)} \quad - (3)$$

Lift:

This indicates how probable item Y is will be purchased when item X is purchased, while controlling for the popularity of item Y. It is computed as shown in equation 4.

If x => y then

$$Lift = \frac{support}{support(x) * support(y)} \quad - (4)$$

The table below explains how to calculate support, confidence, and lift values using the textbook and Score Status.

Table 3: Calculating Support, Confidence, Lift Value Based On Text Book and Score Status

| Subject Name | Text Book | Score_Statu s | Support | Confidence | Lift |
|---------------------|---|------------------|----------|------------|-------|
| Maths-I | N.P.Bali Engineering Mathematics, Lakshmi Publications | positive | 0.007920 | 0.8 | 1.337 |
| Distributed Systems | Distributed Systems Principles Paradigms-Tanenbaum PHI | negative | 0.001980 | 0.5 | 3.556 |
| English-II | English Encounters Published by Maruthi Publishers. | positive | 0.001980 | 0.25 | 1.778 |
| Flat | Principles of compiler design, 2nd edition, Nandhini Prasad, Elsebier | neutral | 0.002980 | 1 | 3.825 |

By considering the textbook and Score Status, we can determine we recommend the strong textbooks based on the 3 parameters. It provides the those textbooks are greater than the minimum support, minimum confidence and minimum lift values. These are calculated as shown above table. In this table contains strong recommended data. Now, we can analyse that some text books who has Score Status is positive have low support, confidence and lift values and some neutral score text books has light support, confidence and lift values. So we can easily remove those conflict textbooks. Because, we only provides the strong textbooks which has positive comments and high ratings.

Finally, we developed a web application by using Django and we provided input details for user want to search. It contains subject name and searching level. This searching level is divided into 2 ways. Those are book level and understanding level. In book level, we provide 3 options like basic level,

intermediate level, advanced level and in understanding level, we provided another 3 options like slow learner, medium-paced learner, and fast learner. Based on the input information provided by the user, this system recommends textbooks and displays comments left by past users.

IV. EXPERIMENT RESULTS:

Here, we are used different algorithms to evaluate the performance of out dataset. This dataset contains 983 records and 10 different attributes. First, we used the label encoding method to transform categorical data to numerical data. Following label encoding, the table looks like this:

Table 4: Data-set For Performance Evaluation By Using Label Encoding

| Sub_tar | Rd_tar | P_tar | No_tar | D_tar | BL_tar | UL_tar | Rat_tar | Score | S_status |
|---------|--------|-------|--------|-------|--------|--------|---------|--------|----------|
| 25 | 2 | 1 | 2 | 1 | 1 | 2 | 4 | 0.4404 | 2 |
| 26 | 2 | 1 | 3 | 2 | 2 | 1 | 3 | -0.053 | 0 |
| 38 | 2 | 1 | 3 | 2 | 0 | 0 | 5 | 0.6249 | 2 |
| 57 | 2 | 0 | 2 | 3 | 1 | 2 | 5 | 0 | 1 |
| | | | | | | | | | |

Here, we have used score status as a class label and this contains negative as 0, neutral as 1, and positive as 2. For evaluating the performance, this data-set is used. After that, the numerical data set is used for performance evaluation. In this evaluation, we are used different algorithms. By using

these algorithms, we are evaluating the accuracies. Each algorithm gives different accuracies. The below algorithms are used for evaluating the performance as follows:

1. Logistic Regression
2. SVM
3. Decision Tree Classifier

The confusion matrix is used to compare the performance of various algorithms. The confusion matrix is a NxN matrix, with N being the number of target classes. The matrix compares the actual goal values to the machine learning models' predictions. The information shown above provides us with a comprehensive perspective of the classification model's performance and allows us to identify the types of mistakes that are created.

The confusion matrix for Logistic Regression is:

```
[[9 2 0]
 [2 30 0]
 [0 0 154]]
```

The confusion matrix for SVM is:

```
[[10 1 0]
 [0 32 0]
 [0 0 154]]
```

In this proposed system, the logistic regression and svm algorithms performs the 3 metrics called as accuracy, precision, and recall.

1. Accuracy : It is determined to measure the best machine learning model. There are various machine learning models. By using them we obtain high accuracy. We identified the relationship and pattern between the

variables of the dataset based on input and training data. It is computed as shown in equation 5.

$$\text{Accuracy} = \frac{\text{Total no. of correct predictions}}{\text{Total no. of samples}} \quad - (5)$$

Accuracy for Logistic Regression : 97.96 %
Accuracy for SVM : 99.49%

2. Precision : Precision is defined as one of the indicators to identify the machine learning model performance. By using the machine learning model, the quality of the precision is made. The quality of precision is based on positive precision and negative precision. True positives are divided by the total number of positive forecasts given. It is computed as shown in equation 6.

$$\text{Precision} = \frac{\text{Sum } i \text{ in } (TruePositives)}{\text{Sum } i \text{ in } (TruePositive + FalsePositive)} \quad - (6)$$

Precision for Logistic Regression : 97.96%
Precision for SVM : 99.49%

3. Recall: Recall is used to identify the number of positives found during the recall session. It is also specified as the correct number of hits has been found. It is computed as shown in equation 7.

$$\text{Recall} = \frac{\text{Sum } i \text{ in } (TruePositives)}{\text{Sum } i \text{ in } (TruePositive + FalseNegative)} \quad - (7)$$

Recall for Logistic Regression : 97.96%

Recall for SVM : 99.49%

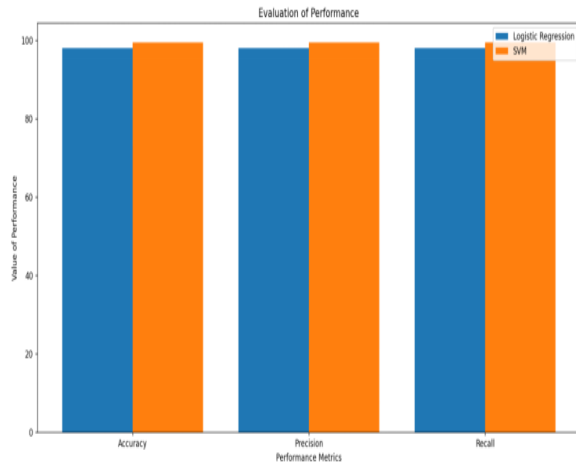


Figure 2: Output Analysis

In these 2 algorithms, SVM gives the best accurate (99.49%) result. We used another algorithm called as Random Forest Regression. It calculates the 3 metrics called as Mean Absolute Error, Mean Squared Error, and Root Mean Squared Error.

1. Mean Absolute Error: The absolute error is the difference between the actual and projected values, whereas the mean absolute error is the weighted average of all absolute errors. It is computed as shown in equation 8.

$$MAE = \frac{\sum_1^n (actual_value - predicted_value)}{n} \quad - (8)$$

Where n is the number of records in the dataset.

The MAE for Random Forest Regression : 0.077

2. Mean Squared Error: It is computed as the mean squared error, which is popular for finding the standard deviation for all the errors that occur during the prediction of class labels by the given model. The corresponding formula for the computation is shown in equation 9.

$$MSE = \frac{\sum_1^n (actual_value - predicted_value)^2}{n} \quad - (9)$$

The MSE for Random Forest Regression : 0.014

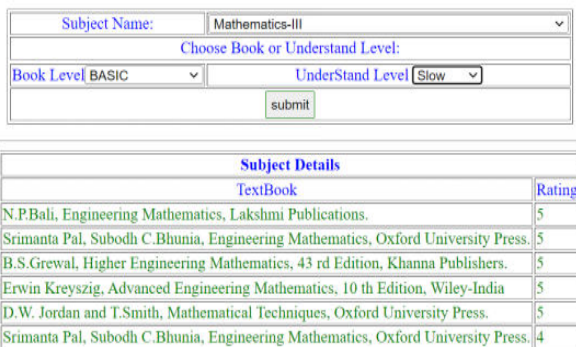
3. Root Mean Squared Error: It is computed as the square root mean squared error, which is popular for finding the standard deviation for all the errors that occur during the prediction of class labels by the given model. The corresponding formula for the computation is shown in equation 10.

$$RMSE = \sqrt{\frac{\sum_1^n (actual_value - predicted_value)^2}{n}} \quad - (10)$$

The RMSE for Random Forest Regression : 0.119

These are the different algorithms are used to evaluated the performance. These algorithms works on the basis of labeled data-set. After this, we developed a web application which performs different operations and it produces the recommended textbooks for the user by taking user specifications. The web application is as sown in below figure 3.

PSCMR Book Recommendation System Form



| Subject Details | |
|--|--------|
| TextBook | Rating |
| N.P.Bali, Engineering Mathematics, Lakshmi Publications. | 5 |
| Srimanta Pal, Subodh C.Bhunia, Engineering Mathematics, Oxford University Press. | 5 |
| B.S.Grewal, Higher Engineering Mathematics, 43 rd Edition, Khanna Publishers. | 5 |
| Erwin Kreyszig, Advanced Engineering Mathematics, 10 th Edition, Wiley-India | 5 |
| D.W. Jordan and T.Smith, Mathematical Techniques, Oxford University Press. | 5 |
| Srimanta Pal, Subodh C.Bhunia, Engineering Mathematics, Oxford University Press. | 4 |

Figure 3 : Web Application Prototype

Table 5: Comparison Table

| Algorithm/Performance metrics | Accuracy | Precision | Recall |
|-------------------------------|----------|-----------|--------|
| Logistic Regression | 97.96 | 97.96 | 97.96 |
| SVM | 99.49 | 99.49 | 99.49 |

| Algorithm/Performance metrics | MAE | MSE | RMSE |
|-------------------------------|-------|-------|-------|
| Random Forest Regression | 0.077 | 0.014 | 0.119 |

V. CONCLUSION:

The primary goal of this suggested approach is to overcome the disadvantages of existing methods for finding an appropriate textbook for a student. In present days also, students are finding the textbook in library in hours of time. In this case, the book recommendation system helps all the users. But all the existing systems have proposed traditional machine learning algorithms but in this proposed system, we used NLP for comment analysis and for producing strong recommended textbooks for the users, we used association rule based apriori algorithm. Before that, for recommending books for new users, we collected the reviews of different books from the previous users. We applied the algorithms to this collected data and finally the recommended textbooks data is produced. For this, we provided the books based on the user searching details. To assess performance, the Logistic Regression and SVM machine learning algorithms are applied. In these algorithms the logistic regression gives 97.96% and svm gives 99.49% of accuracy results. The another algorithm we have used is Random Forest Regression. This algorithm calculates Mean Absolute Error, Mean Square Error, and Root Mean Square Error.

REFERENCES

[1] A. Anoop and N. A. Ubale, "Cloud-Based Collaborative Filtering Algorithm for Library Book Recommendation System," 2020 Third International Conference on Smart Systems and Inventive Technology

(ICSSIT), 2020,pp. 695- 703,
doi:10.1109/ICSSIT48917.2020.9214243.

[2] Dhiman Sarma, Tanni Mitra, and Mohammad Shahadat Hossain, “Personalized Book Recommendation System using Machine Learning Algorithm” International Journal of Advanced Computer Science and Applications(IJACSA), 12(1), 2021, doi: 10.14569/IJACSA.2021.0120126.

[3] Yonghong Tian, Bing Zheng, Yanfang Wang, Yue Zhang, Qi Wu, College Library Personalized Recommendation System Based on Hybrid Recommendation Algorithm, Procedia CIRP, Volume 83, 2019, doi: 10.1016/j.procir.2019.04.126.

[4] Avi Rana and K. Deeba 2019 J. Phys.: Conf. Ser. 1362, doi: 10.1088/1742-6596/1362/1/012130.

[5] Sirikayon, Chalophon & Thusaranon, Panita & Sukpongthai, Warattha. (2018). A preprocessing matrix factorization on collaborative filtering based library book recommendation system. DSIT '18: Proceedings of the 2018 International Conference on Data Science and Information Technology, doi: 10.1145/3239283.3239302.

[6] Abhilasha Sase, Kritika Varun, Sanyukta Rathod, A Proposed Book Recommender System, Volume 4, Issue 2, 2018, doi: 10.17148/IJARCCE.2015.42108.